Aerosol size distribution estimation and associated uncertainty for measurement with a Scanning Mobility Particle Sizer (SMPS)

L. Coquelin[1,2,3,4], N. Fischer[1], C.Motzkus[2], T. Mace[2], F. Gensdarmes[3], L. Le Brusquet[4] and G. Fleury[4]

[1] Département de mathématiques et statistiques, Laboratoire National de Métrologie et d'Essais, 29 avenue Roger Hennequin, 78197 Trappes, France
[2] Département de qualité de l'air et débitmétrie gazeuse, Laboratoire National de Métrologie et d'Essais, 1 rue Gaston Boissier, 75724 Paris, France
[3] Institut de Radioprotection et de Sûreté Nucléaire, BP 68, 91192 Gif-sur-Yvette. France
[4] Supélec Sciences des Systèmes - EA4454 (E3S), 3 rue Joliot Curie, 91192 Gif-sur-Yvette, France

E-mail: Loic.Coquelin@lne.fr

Abstract. Scanning Mobility Particle Sizer (SMPS) is a high resolution nanoparticle sizing system that has long been hailed as the researcher's choice for airborne nanoparticle size characterization for nano applications including nanotechnology research and development. SMPS is widely used as the standard method to measure airborne particle size distributions below 1 μm. It is composed of two devices: a Differential Mobility Analyzer (DMA) selects particle sizes thanks to their electrical mobility and a Condensation Particle Counter (CPC) enlarges particles to make them detectable by common optical counters. System raw data represent the number of particles counted over several classes of mobility diameters. Then, common inversion procedures lead to the estimation of the aerosol size distribution. In this paper, we develop a methodology to compute the uncertainties associated with the estimation of the size distribution when several experiences have been carried out. The requirement to repeat the measure ensures a realistic variability on the simulated data to be generated. The work we present consists in considering both the uncertainties coming from the experimental dispersion and the uncertainties induced by the lack of knowledge on physical phenomena. Experimental dispersion is quantified with the experimental data while the lack of knowledge is modelled via the existing physical theories and the judgements of experts in the field of aerosol science. Thus, running Monte-Carlo simulations give an estimation of the size distribution and its corresponding confidence region.

1. Introduction

Among devices using electrical mobility methods to measure aerosol particle size distribution, the Scanning Mobility Particle Sizer (SMPS) is the most widely used to characterize particles smaller than $1\mu m$. Since it can be applied in fields like: indoors air quality measurement, vehicle exhaust pipes, atmospheric studies, toxicology testing, etc., it has become of interest to assess the uncertainty associated with its outputs. So far, although some works have been carried on the adequacy of particle size measurements made by this device, such a dimensional feature uncertainty cannot be provided. SMPS is composed of a Differential Mobility Analyser (DMA)

and a Condensation Particle Counter (CPC). First one selects particles mobility diameters while the other is a counter used for the detection. First, aerosol passes through an inertial impactor to avoid largest particles to enter the DMA column, then the aerosol enters the charge neutraliser to be conditioned. Indeed, aerosol particles are usually charged. In this case, a bipolar charge distribution is created by exposing the aerosol to a radioactive source, so, particles that carry several charges lose their charge excess. Once aerosol is well conditioned particles are selected using electrical classification inside DMA column: an electric field is created, and the airborne particles drift in the DMA according to their electrical mobility. Once selected, particles reach the CPC to be counted. In this paper, we focus on providing the users with a methodology to compute the uncertainties associated to the size distribution estimated by an inversion procedure. The article is organized as follows: section 2 defines the technical protocol to produce the data, section 3 describes the physical model to relate the experimental data to the size distribution and the inversion procedure to estimate the size distribution. Improvements are mainly brought in section 4 where the sources of uncertainty arising from both the measure and the estimation of the size distribution are decomposed and propagated. Thus, the experimental dispersion is represented by errors on the raw data. Section 4.1 explains how the model of these errors is deduced from some observed signals. Then, section 4.2 describes how the estimation of the size distribution is realized via an inversion model whose variability will refer to the lack of knowledge on the parameters and other physical laws of the system. Preliminary results obtained when this methodology is carried out on SiO_2 airborne nanoparticles are presented in section 5 and it is followed by a discussion about the limits of the methodology and the possible improvements.

2. Technical protocol

Measurements using a SMPS can be performed via several kind of devices. The measurement process can be split into phases: the generation of the aerosol, the neutralization of the charges, the selection of the particles and their detection. In this case, the measurement is assured by a pneumatic atomizer 3076, a charge neutraliser using Krypton source ($^{85}K_r$), a Long-DMA 3081, and a butanol based CPC 3022. Each equipment comes from TSI company. As an example, an other choice could have been an electrospray, a corona charger, a nano-DMA and a water based CPC. The results that will be shown later in this paper refer to the chosen devices.
The protocol that is used is defined as follows:

- Instruments are warmed up for at least 30 minutes
- TSI 3076 operates in the following conditions: air flow rate: 3 L.min^{-1} with inlet pressure at 2.5 kg.cm^{-2}, liquid flow rate: 20 cm^3.min^{-1} at 2.5 kg.cm^{-2} which are nominal settings for the device
- Nanoparticles generated by the atomizer pass through the charge neutralizer 3077, and then, are introduced into the DMA
- DMA settings only consist in defining the following equalities for the flow rates: $Q_{sh} = Q_{exc}$ and $Q_a = Q_{cpc}$ so that the flows are symmetric inside the classifier
- CPC low flow mode is activated such that the inlet flow is 5 cm^3.s^{-1}

Figure 1 presents a schematic diagram of the experimental setup.

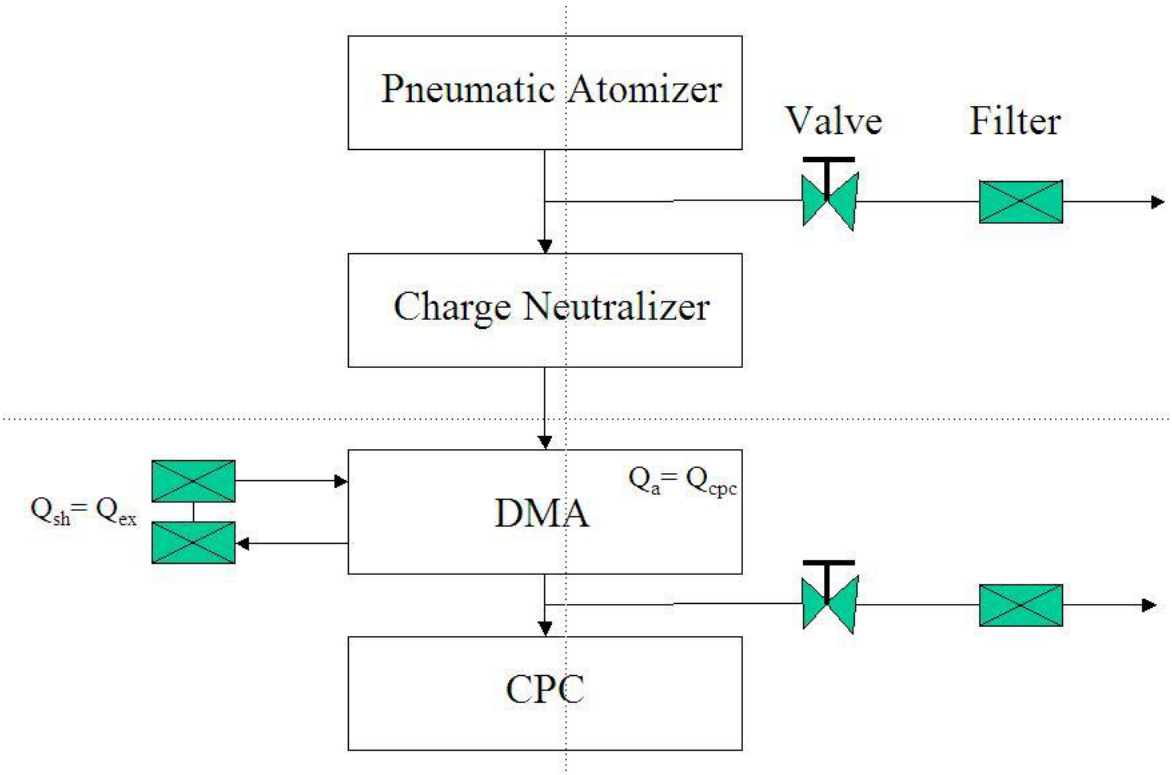

Figure 1. Schematic diagram of the experimental setup

3. Model definition and Inversion procedure

The recovery of size distributions, which is referred as data inversion, requires the solution of the set of Fredholm integral equations. Let y_i be the number of particles counted by the CPC in the ith channel, k_i the nonnegative kernel function for the time range corresponding to the channel i, d_p the mobility diameter of the particle, u the logarithm function, ϵ_i the measurement error in the ith channel and finally f the aerosol size distribution, then:

$$y_i = \int_0^{+\infty} k_i(d_p) f(u(d_p)) d(u(d_p)) + \epsilon_i, \qquad i = 1, ..., m \tag{1}$$

where each kernel function is expressed as:

$$k_i(d_p) = Q_a t_c^i \sum_{p=1}^{+\infty} \Phi(p, d_p) W(d_p) \bar{\Omega}_i(d_p, p). \tag{2}$$

Components Φ, W and $\bar{\Omega}_i$ of the kernel function respectively represent the bipolar charge distribution function, the CPC detection efficiency and the mean transfer function for the time range corresponding to the channel i, p being the number of charges carried by the particle, Q_a the aerosol flow rate entering the DMA and t_c^i the counting time for the ith channel.
The principle of selection is based on the relation between the voltage applied at the center rod of the column of the classifier and the mobility diameter. This relationship is expressed via the electrical mobility. The experimental electrical mobility, denoted as Z_v, can be computed according to the voltage V, the flow rates (Q_{sh} the sheath flow rate entering the DMA and

Q_{ex} the excess flow rate exiting the DMA) and the geometry of the DMA (L the length of the column, r_1 and r_2 respectively the inner and outer radii of the electrodes). The theoretical electrical mobility, noted Z_d, relates Z_v to a specific mobility diameter d_p. Finally considering the equality of both quantities, one can get the mobility diameter associated to the applied voltage.

$$Z_v = \frac{(Q_{sh} + Q_{ex})log(r_2/r_1)}{4\pi LV}, \quad Z_d = \frac{peC_c(d_p)}{3\pi\mu_g d_p}, \tag{3}$$

where e is the charge of the electron, μ_g the dynamic gas viscosity and C_c the slip correction factor. The expression of the slip correction factor is based on the triplet (α, β, γ) and its expression is as follows:

$$C_c(d_p) = 1 + K_n(d_p)\left[\alpha + \beta \exp(-\gamma/K_n(d_p))\right], \tag{4}$$
$$K_n(d_p) = (2\lambda_m)/d_p, \tag{5}$$

where K_n is called the Knudsen number and λ_m is the mean free path of a particle.

Since the DMA voltage is being continuously scanned, the times of the system become crucial. Indeed, particles are selected by the DMA and after transport through the tubing (transport time is denoted as t_d), the counter detects them. An error in the time specification corresponding to the time for particles to pass from the DMA exit slit to the CPC counting chamber lead to an error in the estimation of f on the diameter space. The same kind of statement can be made in what concerns the time for particles to pass through the DMA column (denoted as t_f). This time is computed as the time for the aerosol to pass through the DMA.
The discrete form of equation 1 after applying a numerical quadrature to approach the integral numerically can be written as:

$$y_i = \sum_{j=1}^{j=n} w_{ij} k_{ij} f_j + \epsilon_i, \quad i = 1, ..., m, \tag{6}$$

where $k_{ij} = k_i(d_{p,j})$, $f_j = f(d_{p,j})$, $(d_{p,j})_{j=1,...,n}$ being the vector of points used for the discretization of the diameter space. Once the relationship between the inputs and the outputs of the measurement system is well-defined, next step is to estimate each component of the vector f. Several algorithms lead to comparable estimations of the size distribution. Most commonly used algorithms in the field of aerosol are those developed by Seinfeld [1], Twomey [2],[3] and more recently Collins *et al.* [4] or Talukdar [5]. Following Seinfeld's method, inversion with regularization techniques has been chosen here. It leads to a fast and accurate estimation of f and it also takes into account the judgements of the experts through the prior. Let H the matrix such that $H = (w_{ij}k_{ij})_{i=1,...,m}^{j=1,...,n} \in R^{m\times n}$, $y = (y_i)_{i=1,...,m} \in R^m$ and $f = (f_j)_{j=1,...,n} \in R^n$, then the inversion procedure becomes the minimization problem: find $f^* \in R^n$ such that

$$f^* = \arg\min_{f\geq 0} \left\{ \|Hf - y\|_2^2 + \lambda \|D_2 f\|_2^2 \right\}, \tag{7}$$
$$= \arg\min_{f\geq 0} \left\{ G\left(H, f, y, \lambda\right) \right\}, \tag{8}$$

where D_2 is the second-order difference matrix and $f \geq 0$ means $f_j \geq 0$, $\forall j$.

Here the prior penalizes solutions that are not smooth. This type of penalty seems justified in this application because aerosol processes often tend to smooth rough distributions. The

parameter λ will be a trade off between the prior we associate to the solution to be estimated and the solution that minimizes the quadratic error term. The L-curve method is chosen here to determine the regularization parameter and it is computed via the pruning algorithm developed in [6]. To ensure robustness, λ is calibrated through the all set of data. The solution of the minimization is computed via the fast non negative least-square algorithm (*fnnls* [7]) algorithm that is an optimized version of Lawson and Hanson non negative least-square algorithm (*nnls* [8]). Several references [1],[2],[3],[4],[5] define how to estimate the size distribution, this article does not intend to bring improvements in the inversion procedure but mainly in the following section dedicated to the uncertainty analysis.

4. Uncertainty analysis

In this part, the aim is to quantify and to propagate the uncertainties coming from both the measurement process and the inversion. The experimental dispersion is difficult to model since each component of uncertainty that it involves is not accurately modelled yet. Thus, choice has been made to quantify it with experimental data. It can be seen as a behavioural model and it has the advantage that it warranties a realistic quantification of this source. On the other hand, the uncertainty associated with the inversion procedure is modelled. Unlike for the experimental dispersion, the uncertainty it induces cannot be track down into the data because the error it produces is systematic. The major source is the lack of knowledge on the physical models to be used to define the matrix H. Indeed, several theories compete to define the factors of Cunningham, the bipolar charging law or the diffusive transfer function. Moreover, elements of this group are different in terms of nature: scalars and functions. Some scalars can be modelled with corresponding uncertainty because they can be measured (L, r_1, r_2). Some cannot be measured but the existing theories can be used to bound them (α, β, γ) or concerning the functions ϕ and Ω, comparisons between data and models have been performed under controlled conditions. In this paper, we will only consider the ignorance on parameters, so Fuch's theory [9] is chosen to define ϕ and Stolzenburg's work [10] is used to model the diffusional transfer function. To sum up, the experimental dispersion, denoted as the first group of uncertainty represents 3 major components:

- the instability of the aerosol generation,
- the partial variability of the physical parameters (flow rates, environmental conditions, etc.),
- the noise coming from the CPC counting process,

while the inversion uncertainty that is considered as the second group is made of:

- the lack of knowledge associated with the choice of the physical parameters,
- the variability induced by the choice of the regularization parameter,
- the bias due to the method of regularization.

The evaluation of the bias requires an aerosol of reference from a metrological point of view. Since it is not available due to the instabilities of the generation process, it is not yet possible to account for it in the uncertainty budget.

4.1. The experimental dispersion

In practice, a set of measurements have been made and the user's got observations $(y_{il})_{i=1,\dots,m}^{l=1,\dots,n_l}$, i being the index for the class of diameter and l the index that corresponds to the number of experiences. We consider that the experiences have been repeated n_l times, so that enough information can be extracted from the data. The idea we propose is to infer a model based on this information by considering simulated data modelled as a correlated Gaussian process noted $\tilde{y}$. Experiences tend to justify the Gaussian assumption and authors usually abound

in that sense. Let C be the empirical correlation matrix, the autocorrelated Gaussian kernel is computed by performing a SVD (Singular Value Decomposition) on C such that it can be decomposed as $C = U\Sigma V^*$.
$\tilde{y}$ is then created as follows:

$$\tilde{y} = \left(U\sqrt{\Sigma}z\right)\sigma + \mu, \tag{9}$$

where $z = (z_1, z_2, ..., z_m)$ with $z_i \sim N(0,1)$, μ and σ are respectively the mean and standard deviation of the real observations: $\mu = (\mu_1, \mu_2, ..., \mu_m) \in R^m$ such that $\mu_i = \frac{1}{n_l}\sum_{l=1}^{l=n_l} y_{il}$, $\sigma = (\sigma_1, \sigma_2, ..., \sigma_m) \in R^m$ such that $\sigma_i = \frac{1}{n_l - 1}\sum_{l=1}^{l=n_l}\sqrt{(y_{il} - \mu_i)^2}$. To propagate the uncertainty arising from the experimental dispersion, the methodology is to generate a set of simulated data via Monte-Carlo simulations and to estimate the vector f via the minimization described in (eq 7). Let $\tilde{y}^{[q]} \in R^m$ be the simulated data generated for the q^{th} simulation, n_q the total number of simulations and $\bar{\lambda}^{Lcurve}$ the regularization parameter calibrated thanks to the set of real data y via the L-curve algorithm, then the estimation of the solution vector f for the q^{th} simulation denoted as $\hat{f}^{[q]} \in R^n$ is computed as:

$$\hat{f}^{[q]} = \arg\min_{f \geq 0}\left\{G\left(H, f, \tilde{y}^{[q]}, \bar{\lambda}^{Lcurve}\right)\right\}, \quad q = 1, ..., n_q. \tag{10}$$

4.2. The inversion uncertainty
A parametric model is developed in order to take into account the uncertainty associated with the lack of knowledge on parameters to be used for the definition of the matrix H. These parameters are selected such that they are not well-known or because several theories compete to define them and especially because they have a significant effect on the final estimation. In a previous paper [11], sensitivity analysis has been carried out to determine the most significant parameters, significant means here that a small error in their definition produces a large variability on the reconstruction of the solution vector. The chosen parameters are: L, r_1, r_2, α, β and γ. Nominal values for α, β and γ are taken from experiments made by Kim *et al.*. [12]. Experiences realized by Hutchins *et al.*. [13] and Allen and Raabe [14] are chosen as bounds around the nominal values. Subscripts $_k$, $_h$ and $_{ar}$ refer to each author. Nominal values for the geometrical parameters of the DMA are given by the constructor and their associated uncertainty is taken from the ISO standard (reference ISO 15900:2009(E)). To account for the uncertainty the selected parameters produce, they are modelled as random variables and propagate through the inversion via Monte-Carlo simulations as for the experimental dispersion. Table 1 summarizes the choices made for the definition of the random variables. Let $\tilde{H}^{[q]}$ be the matrix drawn by the random variables for the q^{th} simulation, for a chosen experimental measurement $y = (y_{il})_{i=1,...,m}^{l=l'}$, $\hat{f}^{[q]}$ is now computed as:

$$\hat{f}^{[q]} = \arg\min_{f \geq 0}\left\{G\left(\tilde{H}^{[q]}, f, y, \bar{\lambda}^{Lcurve}\right)\right\}, \quad q = 1, ..., n_q. \tag{11}$$

Table 1. Definition of the parameters that model the lack of knowledge. U and N respectively refer to Uniform and Normal distributions. Subscript $_0$ refers to the nominal value given by the constructor for the corresponding parameter.

Name	Law	Mean	Standard deviation	Min	Max
r_1	U	$r_{1,0}$	-	$r_{1,0} - 0.001 r_{1,0}$	$r_{1,0} + 0.001 r_{1,0}$
r_2	U	$r_{2,0}$	-	$r_{2,0} - 0.0006 r_{2,0}$	$r_{2,0} + 0.0006 r_{2,0}$
L	U	L_0	-	$L_0 - 0.005 L_0$	$L_0 + 0.005 L_0$
α	N	α_k	$\frac{\max(\|\alpha_h - \alpha_k\|, \|\alpha_{ar} - \alpha_k\|)}{\alpha_k}$	-	-
β	N	β_k	$\frac{\max(\|\beta_h - \beta_k\|, \|\beta_{ar} - \beta_k\|)}{\beta_k}$	-	-
γ	N	γ_k	$\frac{\max(\|\gamma_h - \gamma_k\|, \|\gamma_{ar} - \gamma_k\|)}{\gamma_k}$	-	-

The decomposition of the sources of uncertainty as presented above is required when the purpose is to characterize their respective effect on the reconstruction of the solution vector f. Nevertheless, the final result is the estimation of f with the associated uncertainty and it has to be computed when both sources are simultaneously considered. Combining the sources is straightforward since the procedure is the same as for each source taken individually. Thus, when both sources are combined, the estimation of the solution vector f for the q^{th} simulation is obtained as:

$$\hat{f}^{[q]} = \arg\min_{f \geq 0} \left\{ G\left(\tilde{H}^{[q]}, f, \tilde{y}^{[q]}, \lambda^{[q],Lcurve} \right) \right\}, \quad q = 1, ..., n_q, \tag{12}$$

where $\lambda^{[q],Lcurve}$ is the regularization parameter computed for the q^{th} simulation.

5. Results

Starting from real observations, the aim of this part is to show the uncertainty associated with the mean estimate, denoted $\bar{\hat{f}}$, that is computed with the methodology presented above for each source taken individually and combined. As an example, SiO_2 airborne nanoparticles are presented. The aerosol presents 2 populations in terms of size: the first one is around 30 nm while the second population stands around 40 nm. Dispersions around the peaks are similar. These observations will prove the robustness of the inversion procedure, indeed, both peaks are really close and a bad estimation of the regularization parameter shall lead to over-smooth the solution. Figure 2 shows the real observations y ($l = 40$ in this case) and the corresponding simulated data generated via Monte-Carlo simulations. Following the developments proposed in section 4, experimental dispersion and its effects on the estimation of the vector f as well as the lack of knowledge and its impact on the reconstruction are illustrated in figure 3. Combined uncertainties of both sources are also presented in this figure. The uncertainties are shown through a confidence region with a confidence level of 95 %. The results suggest that the experimental dispersion affects the estimation of f on the y axis (concentration of particles) with the same amplitude through the all diameters while the lack of knowledge on the physical phenomena does not significantly affect the concentration of particles but it induces a translation

of the reconstruction. Other experiences are actually being realized on different aerosols in order to confirm these results.

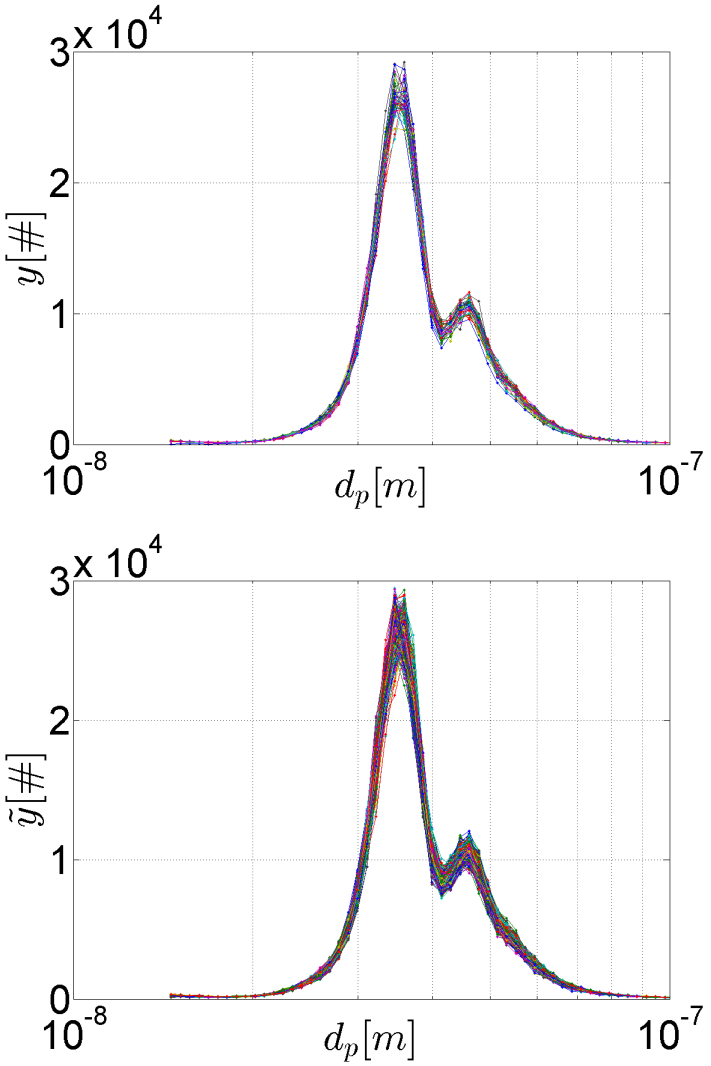

Figure 2. Real data (top) versus simulated data (bottom) for SiO_2 airborne nanoparticles

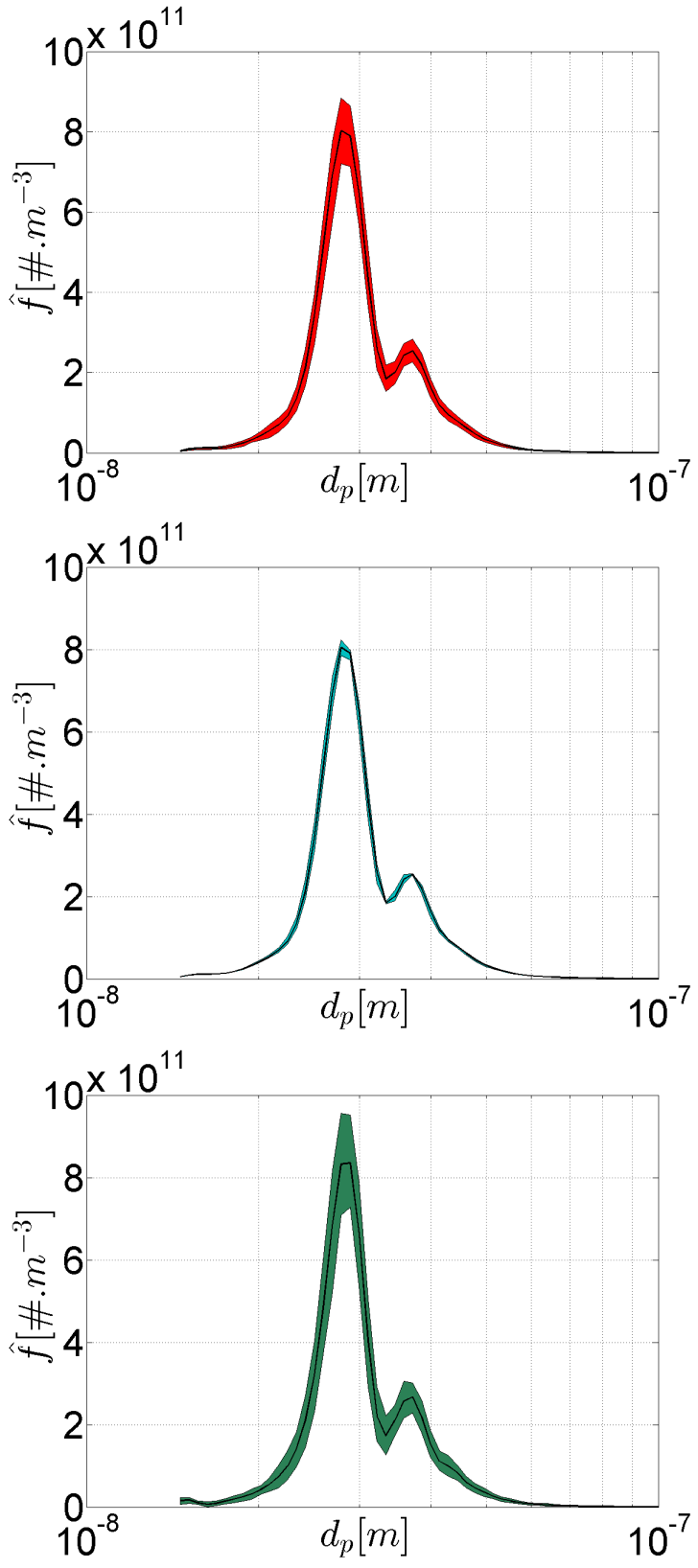

Figure 3. Illustrations of the 95 % confidence region computed thanks to Monte-Carlo simulations for each source of uncertainty. The figures show respectively the uncertainty associated with the experimental dispersion (top), the lack of knowledge (middle) and both sources combined (bottom)

6. Conclusion

A methodology has been presented to compute the confidence region on the estimated size distribution when several experiences have been carried out to characterize an aerosol. The statistical model that accounts for the experimental dispersion has been inferred from real observations because spatial correlations are difficult to model precisely. The statistical inversion procedure takes into account the uncertainties coming from the lack of knowledge on parameters that are used to relate the data to the size distribution through the matrix of kernel functions. An example of aerosol composed of SiO_2 airborne nanoparticles is treated using this methodology. Results reveal that the uncertainty arising from each source affects respectively the reconstruction on the concentration of particles and on the diameters. Some sources still need to be added especially the impact of the prior in the minimization problem and the bias that it introduces on the estimation of f. Next step will be to consider analytical size distributions to estimate the bias induced by the prior since an aerosol of reference is not yet available. An other improvement would be to precisely model and propagate the uncertainties when only one experience is available.

References

[1] Seinfeld J H and Wolfenbarger J K 1990 *Journal of Aerosol Science,* **21** 227–247

[2] Twomey S 1963 *Journal of the Association for Computing Machinery* **10** 97–101

[3] Twomey S 1975 *Journal of Computational Physics* **18** 188–200

[4] Collins D R, Flagan R and Seinfeld J H 2002 *Aerosol Science and Technology* **36** 1–9

[5] Talukdar S S and Swihart M T 2003 *Aerosol Science and Technology* **37** 145–161

[6] Hansen P C and O'Leary D P 1993 *SIAM Journal on Scientific Computing* **14** 1487–1503

[7] Bro R 1997 *Journal of Chemometrics* **11** 393–401

[8] Lawson C L and Hanson R J 1995 *Society for Industrial and Applied Mathematics* **15** 345–349

[9] Fuchs N A 1963 *Geofis. Pura Appl* **56** 185–193

[10] Stolzenburg M R 1988 *An Ultrafine Aerosol Size Distribution Measuring System* Ph.D. thesis University of Minnesota, Minneapolis

[11] Coquelin L, Fischer N, Le Brusquet L, Fleury G, Motzkus C and Gensdarmes F 2012 *Instrumentation and Measurement Technology Conference (I2MTC)* (2012 IEEE International) pp 1718–1721

[12] Kim J H, Mulholland G W, Kukuck S R and Pui D Y H 2005 *Journal of Research of National Institute of Standards and Technology* **110** 31–54

[13] Hutchins D K, Harper M H and Felder R L 1995 *Aerosol Science and Technology* **22** 202–218

[14] Allen M D and Raabe O G 1985 *Journal of Aerosol Science* **4** 269–285